AUTOMATIC CLASSIFICATION OF VOICE DISORDERS IN COURSE OF NEURODEGENERATIVE DISEASE

T. Orzechowski[1], A. Izworski[1], I. Gatkowska[2], M. Rudzińska[3]

[1]Department of Automatics, AGH University of Science and Technology, Krakow, Poland
[2]Computer Linguistic, Jagiellonian University, Krakow, Poland
[3]Collegium Medicum, Jagiellonian University, Krakow, Poland

I. INTRODUCTION

The study presented in this publication is the first from the planned complex, interdisciplinary studies. The examination was carried out on patients of CM-UJ clinic in Krakow who suffered from neurodegenerative disease with the damage of the extrapiramidal system with dysarthria-type changes in speech. Control examinations of healthy persons have also been carried out. The elements whose realization was tested had been chosen based on the linguistic knowledge in the scope of phonetics as well as on experience resulting from long-term practice as a speech pathologist. The linguistic material was selected in such a way as to pinpoint voice changes characteristic for patients with dysarthria. During the examination, phrases based on Polish idioms were aslo recorded for further analyses.

II. VOICE PHYSIOLOGY

Voice and speech production requires close cooperation of numerous organs which from the phoniatric point of view may be divided into organs:

- producing expiration air stream necessary for phonation (lungs, bronchi, trachea),
- amplifying the initaial tone (larynx),
- forming tone quality and forming speech sounds (root of the tongue, throat, nasal cavity, oral cavity).

III. VOICE PATHOLOGY

Apart from typical changes caused by neurodegenarative disease (e.g. shivering of the body, limbs, muscle stiffness) changes in the voice may also be observed. The research shown in work (Intensive voice treatment LSVT® 2001) indicate the serious problem of speech pathology occurance with as much as 75% of patients. Thus it may be concluded that voice constitutes one of the more crucial components of neurological diagnosis.

Patients suffering from neurodegenerative diseases (and such patients were examined by the authors) show dysarthria-type speech alternations. Dysarthria is a group phonation and articulation disorder which result from damage to the movement control systems of the central or peripheral nervous system also responsible for the speech apparatus. The disorders occur although the speech plan is preserved [3]. Other definitions characterize dysarthria as handicapped production of articulated speech sounds resulting from disturbances to nervous mechanisms of voice production, modulation, intensity, timbre and resonance [2]. Nowadays dysarthria is described as a group of motor speech impairment result from a disruption of muscular control due to lesions of either the central or peripheral, or both, nervous systems. Communication Independence for the Neurologically Impaired CINI – 1994).

Due to the dominating symptom of disorder [3] 6 types of dysarthria have been specified. In our study, patients suffered from hypokinetic and hyperkinteic types. Parkinson disease and Parkinson syndrome (damage to the extrapramidal system; speech impairment related to slowness) are accompanied by hypokinetic dysarthria-type changes in speech. Its most important characteristics in relation to isolated sounds are: distortions, loudness limitations. Distorted articulation is caused by quick and limited tongue and lips movements, sounds reduced down to slurring. Impairment in the speech process consist in sudden pauses in phonation. The voice is monotonous, quiet, weak and vanishing. The other type of dysarthria occurring in neurodegenerative diseases of the extrapyramidal system is hyperkinetic dysarthria. Phonation is distorted, sudden pauses in speech may occur. Moreover, incorrect articulation occurs as well as irregular breaks in articulation, sound elongation, repetition of sounds caused by abnormal muscular tension. Hypernasality may also occur, and the loss of air caused by throat and palate impairment result in the shortening of phrases. There are variations of speech loudness, the voice is trembling, tense and stifled, weak, with breaks.

IV. CHARACTERIZATION AND CLASSIFICATION OF SOUNDS USED IN THE EXAMINATION

During the examination both consonants and vowels were used. Patients were asked to pronounce the sounds in isolation.

Models and analysis of vocal emissions for biomedical applications. 4th international workshop. October 29-31, 2005 – Firenze, Italy.
Edited by C. Manfredi. ISBN 88-8453-320-1 (online)

The vowel group consisted of [a], [e] and [i]. This particular choice was related to the difference in the elevation of the tongue as well as to the gap between the lips.

	Division related to the degree of tongue elevation	Division related to the degree of mouth opening
[i]	high	ajar
[e]	medium-high	half ajar
[a]	low	open

Tab.1. Division of vowels.

The closer the tongue to the hard or soft palate, the smaller the degree of the oral gap, the lower the tongue, the bigger the gap. The position of the tongue in relation to the horizontal axis of the oral tract constitutes the basis for the division of vowels into more or less front or back. In our examination we used front vowels. High front vowel [i] is characterized by the very close position of the middle part of the tongue moving up the oral cavity towards the hard palate. In the case of the low front vowel [a], both the hump on the low-situated tongue as well as the spot on the harb palate, towards which the tongue rises, are situated a bit more to the back.

The consonant group consisted of [s], [x], [p], [k] and [g]. This category of sounds may also be divided into groups and subgroups, based on various articulation criteria. One of them is the manner of articulation, limiting to a different degree the flow of air throught he voice channel, up to a complete lack of flow.

[s] and [x] sounds are examples of fricatives. They are consonants in the articulation of which particular parts of the speech organ move closer together creating a narrow gap. Airflow which has proper mass and speed passes through the gap and is disturbed. This gap may be formed in various places of the vocal tract under the larynx. The [s] consonant belongs to front-tongue dental speech sounds, whereas [x] belongs to back-tongue palatal speech sounds.

Sounds [p], [k] and [g] are plosives. The first phase of their duration consists in a solid obstruction built up somewhere within the oral tract, initially completely blocking the airstream coming up from the larynx. This blockage is then usually released abruptly, so that the air that was compressed behind the obstacle can escape with a kind of explosive movement, producing a ‘cracking’ or ‘popping’ sound.

The [p] consonant is a bilabial, whereas [k] and [g] are back-tongue palatal consonants.

5. EXAMINATION METHOD

The examinated group consisted of 18 patients between the ages of 20 and 80 and a comparative group of healthy persons with similar age range. Patients suffered from hypokinetic and hyperkinetic types of movement disorders. The voice of the examined patients was recorded with high-qiality digital equipment in a soundproof room in order to eliminate any undesirable factors which could negatively affect the results. First, particular sounds were isolated from the recorded voice and then they were processed (filtration and spectrum analysis). Spectrum analysis contains numerous details, thus parameterization was necessary for automatic classification.

Firstly, the features of the spectrums of diagnostically essential sounds were verified.

V.1. Changes in sounds realization

Voice signals consist of several waves with different frequencies and amplitudes. The inner ear of humans decomposes the incoming acoustical waves into separate frequencies. Thus, it is appropriate to transform the PCM data into the frequency domain before analyzing it further. This can be achieved using Fourier Transformations.

Using the linear Fourier transform, a continuous signal can be transformed between its time domain representation, denoted by $h(t)$, and the frequency domain representation $H(f)$.

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-j2\Pi ft}dx \qquad (1)$$

The audio signal is sampled at a fixed sampling rate, so the function is not continuous $h(t)$ but discrete x(k).

Consider a series x(k) with N samples of the form x_0, x_1, … ,x_{N-1}

$$X(n) = \frac{1}{N}\sum_{k=0}^{N-1} x(k)e^{-jk2\Pi n/N} \, for \;\; n = 0..N-1 \qquad (2)$$

If N is a power of 2, the Fourier transform can be calculated very efficiently. It is known as *Fast Fourier Transformation* (FFT), and implemented in most of languages of technical computing (e.g. MATLAB®).

The power spectrum matrix P($n; t$), where n is the index for the frequency and t for the time frame:

$$P(n,t) = \left|X_t(n)\right|^2 \frac{1}{N} \qquad (3)$$

The index n ranges from 1 to N=2+1.

It is convenient to use the Bark Frequency Scale instead of Hz. The name has been chosen in memory of Barkhausen, a scientist who introduced the phon to describe loudness levels for which critical-bands play an important role. The Bark scale ranges from 1 to 24 Barks, corresponding to the first 24 critical bands of hearing [Hz].

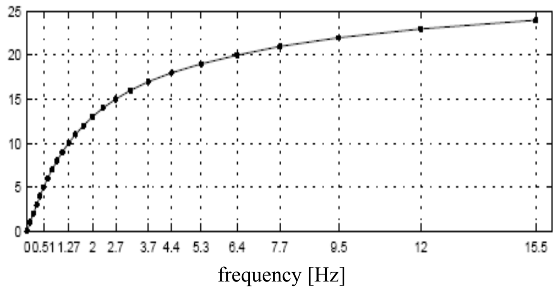

Fig.1. Charakteristic of Bark Frequency Scale

A critical-band value is calculated by summing up the values of the power spectrum within the respective $f_{low}(i)$ and $f_{hig}h(i)$ frequency limits of the i critical-band.

$$CB = \sum_{n \in I(i)} P(n,t)$$
$$I(i) = \{n : f_{low}(i) < f(n) \leq f_{high}(i)\} \qquad (4)$$

where i, t, n are indexes, CB is a matrix containing the power within the i-th criticalband at a specific time interval t.

With the patients, changes in sounds articulation are visible (precisely, transition into another sound during realization). It is both audible and detectable through spectrum comparison. These changes were particularly observable for the following consonants, for which the occurring change has also been indicated:

- [k] → [a] / [k] → [y]
- [g] → [y] / [g] → [e]
- [s] → [y]
- [h] → [a]

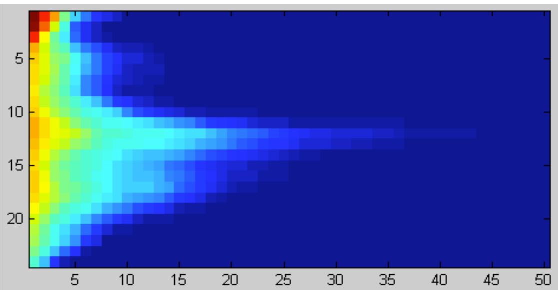

Fig.2a. Voice signal [k] without changes

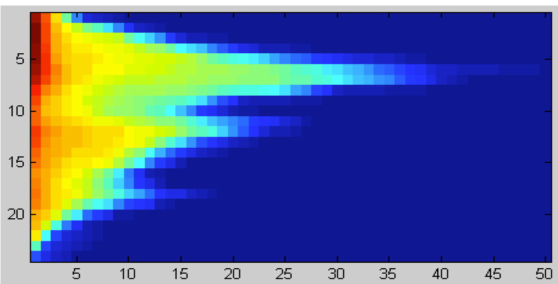

Fig.2b. Voice signal [k] with changes [k] → [a]

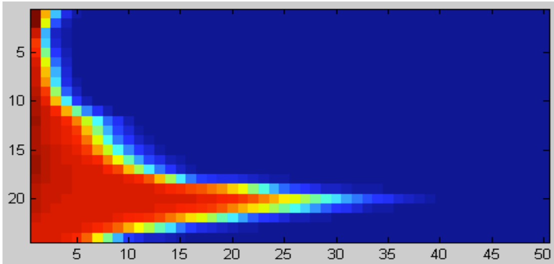

Fig.3a. Voice signal [s] without changes

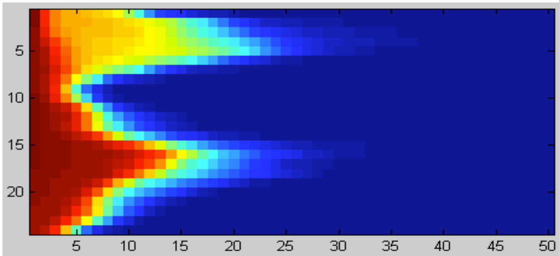

Fig.3b. Voice signal [s] with changes [s] → [y]

Each piece of voice is represented by CB matrix. Firstly, the information represented each group of people was combined using median method. The median proved to be the simplest approach with a comparable quality to other more complex methods. Classification was done using simple distance comparation between CB matrixes. This distance can be used as another voice characterization parameter.

Ther result of this distortion is caused by the weakening of the elasticity of the larynx muscles, that is why a consonant is followed by a vowel, which does not require as much tension.

V.2. Intensity of sounds pronouced many times in isolation

The patients were asked to repeat the same plosive four or more times. The request was based on the knowledge that during the realization of a sequence of the same sounds, sound sequence distortions could be expected and slurring could occur. By analysing the duration and intensity of consecutive sounds, it was noticed with most of the patients that the intensity of the last sounds was respectively lower than the intensity of the first sound. This is due to muscle stiffness, characteristic for Prakinson disease. These sypmtoms were not observed with most of the members of the control group.

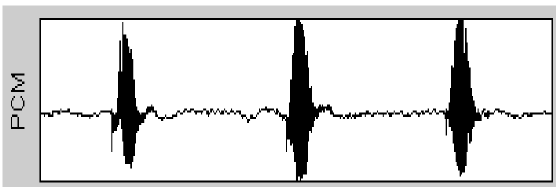

Fig.4a. Last 3 sounds of the healthy person, [p]

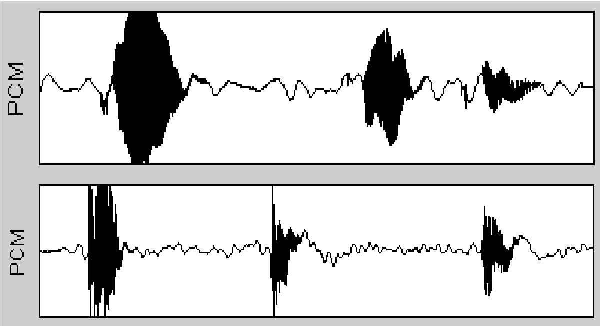

Fig.4b. Last 3 sounds of patients with neurodegenerative diseas.

V.3. Continuous sound analysis

The patients were asked to pronounce the tested sounds [a], [e], [i], [s] or [x] on one breath. The sound emitted for a long period of time allowed for spectral analysis aiming at observing the transition of frequency changes related to pathological trembling. The values received were compared with the values obtained in the control group. With some patients, a slight difference in the voice spectrum in the range between 4-8 [Hz] was observed. This range is characteristic for Prakinson disease tremor. However, at this stage of the study, the results are not reliable enough and require further work, in order for this element to be another element as another voice characterization parameter.

With many patients, distinct and varying breaks in phonation were observed. With healthy persons, gradual quietening took occurred, whereas the patients ended the emission abruptly.

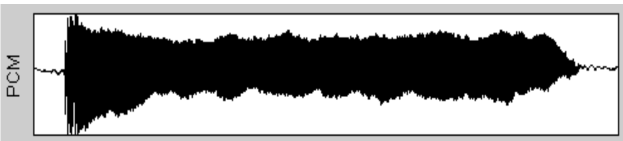

Fig.5a. PCM of healthy person, continuous [a]

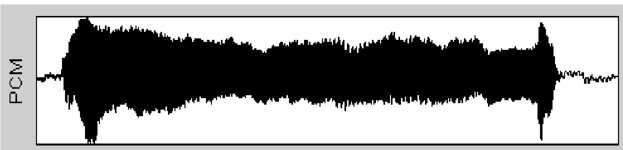

Fig.5b. PCM of person with neurodegenerative diseas.

The authors analyse differences within the range of realization of vowels at different hights.

6. Conclusion

The results presented here constitute the beginning of tests concentrated on automatic voice classification. The authors refered both to the question of duration parameterization as well as voice spectrum parameterization. The main goal set for the future sudies is such defining of deecriptors, which together with particular search algorithms will enable proper interpretation of a patient's voice changes. The proposition of recording, processing and analysis of speech as a digital signal is also presented.

Further analysis of the isolated sounds is planned, compared in realization between the patients and healthy persons, taking sex, age and phrase analysis into consideration. With patients the dynamics of disease progression is also registered.

Moreover, some linguistic material on the level of phrases was recorded and a technical analysis is being prepared. In this study, the prosodic elements of speech – rhythm, pace, intonation, accent and melody will be analysed. The elements mentioned above are available in subjective diagnostic (defining the type of dysarthria). The authors wish to examine the characterization of changes in speech unavailable in subjective examination as well as to create a complex model of automatic classification. In order to achieve this, the newest methods of signal recording, processing and analysis will be implemented.

Bibliography

[1] P. Duus: *Neurologisch-Topische Diagnostik,* Stuttgart, 1983.

[2] D. F. Johns: *Clinical management of neurogenic communicative disorders*, Boston, College Hill, 1985

[3] F. Darley, A. Aronson, J. Bron: Cluster of deviant speech dimension in the dysartrias, *Journal of Speech and Hearing Research*, 12, pp. 462-496, 1969

[4] L. Ramig, S. Sapir, S. Countryman: Intensive voice treatment (LSVT®) for patients with Parkinson's disease: a 2 year follow up., *J Neurol Neurosurg Psychiatry*,7, 1, pp.439 – 498, 2001

[5] H. Wróbel: *Gramatyka współczesnego języka polskiego - Fonetyka i fonologia,* Kraków, PAN, 1995 (in Polish).

[6] A. Pruszewicz: *Foniatria kliniczna,* PZWL, Warszawa, 1992 (in Polish).

[7] A. Izworski, R. Tadeusiewicz: Artificial Intelligence Methods in Diagnostics of the Pathological Speech Signals, in: *Speech and Language Technology*, vol. 6, PPA, Poznań, pp. 183 – 197, 2002

[8] E. Pampalk: *Islands of Music, Analysis, Organization, and Visualization of Music Archives,* Technischen Universitat Wien, 2001.

[9] A. Ralston: *A first course in numerical analysis*, New York, 1965.

[10] T. Körner: *Fourier Analysis*, England, Cambridge University Press, 1988.

[11] D. C. Hanselman, B. L. Littlefield: *Mastering MATLAB 7,* 2004